



Managing Research Data

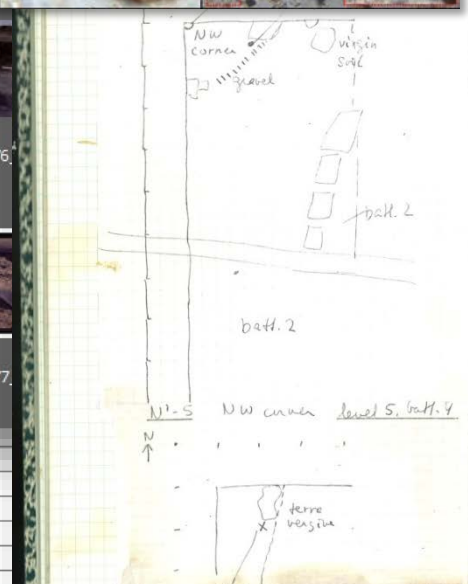
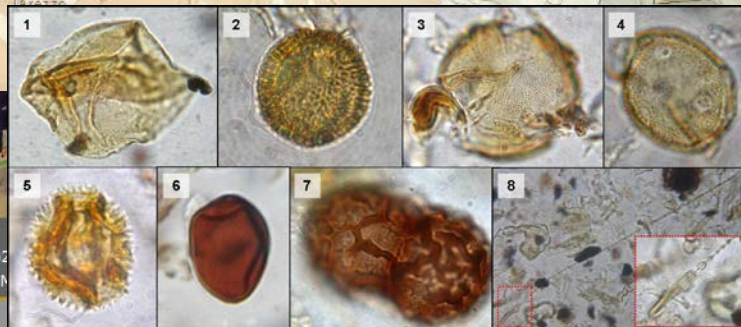
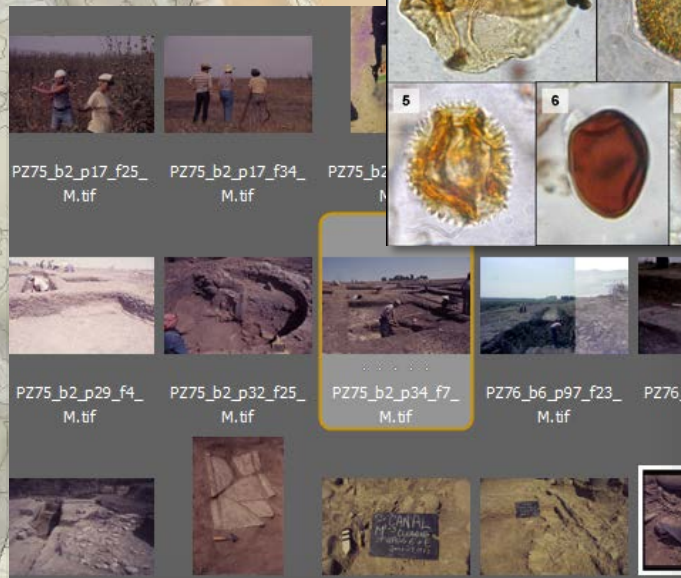
Jessica Trelogan

Data Management Coordinator,
UT Libraries

j.trelogan@austin.utexas.edu



What are "data"?



PIZZICA, 1977 TUESDAY, 5 JULY, EDLUND

N1-S, 01-S Level 5, bath. 4

purpose: to define extent of virgin soil (which appeared in N.E. corner of bath. 3) and black soil (N.W. corner)

black heavy soil - lighter shade in between (irregular to that of bath. 2)

PZ 77-743 P PZ 77-973 P

PZ 77-724 P finds: few shreds clay and to bath 3

PZ 77-728 P NW corner: miniature vase

PZ 77-725 P Complete TC female figurine pink now brown 20 cm. (lost from stake in N1-S)

at the end of bath. 4, three vases appeared in a N-S line with black soil on either side - the surface was cleaned and photos graphed.

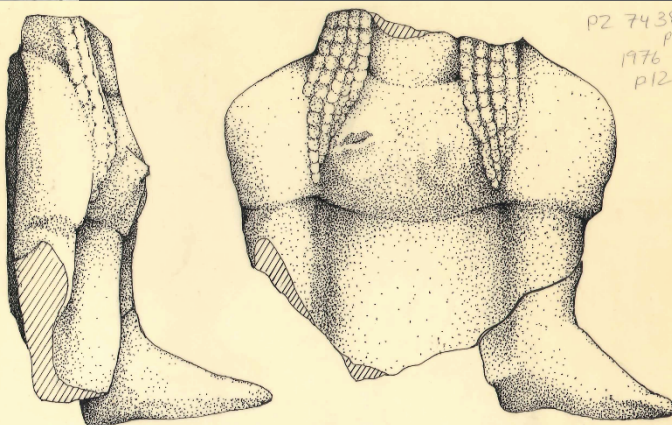
level 5, bath. 5

again concentration of good sh. cent. material in NW corner - miniature vases, wine and house of obols.

01-S level 5, bath. 3 3 end, E of well, PZ 77-965 P

finds: bronze shagil PZ 77-735 M

PZ 77-966 P, PZ 77-967 P, PZ 77-968 P



Harrowing	35	65	Small cemetery	Light	90	2007	Light
None	75	25	Other agrarian	Light	90	2007	Heavy
Harrowing	20	40	Farmhouse	Very light	90	2007	Light
Harrowing	5	70	Other agrarian	Very light	90	2008	Moderate
Harrowing	35	65	Other agrarian	Light	90	2007	Light
None	50	50	Other agrarian	Heavy	85	2005	Heavy

What are “data”?

Natural/Physical Sciences

Observational

Experimental

Simulation

Compiled

Social Sciences

Qualitative

Quantitative

Humanities

Raw

Primary

Interpretive/Derived

National Science Foundation:

“...determined by the community of interest through the process of peer review and program management. This may include, but is not limited to: **data**, publications, samples, **physical collections**, software and models.”

National Endowment for the Humanities

“...materials generated or collected during the course of conducting research.”

Includes:

- citations
- software code
- algorithms
- digital tools
- documentation
- databases
- geospatial coordinates
- reports and articles

Excludes:

- preliminary analyses
- drafts of papers
- plans for future research
- peer review assessments
- communications
- confidential materials
- information violating privacy

What is Data Management?



condensedconcepts.blogspot.com/2009_09_01_archive.html

A collection of tasks practiced throughout the lifecycle of research that make it easier to find, understand, navigate, and use your data.

Why bother?



"I was close to a breakthrough when the grant money ran out."

Reprinted from Funny Times / PO Box 18530 / Cleveland Hts. OH 44118
phone: 216.371.8600 / email: ft@funnytimes.com

Save time and money

Maximize your impact

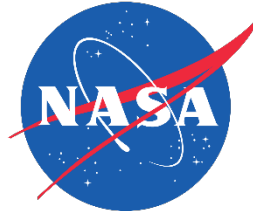
Allow for reuse

Do better research

Why else?



NATIONAL ENDOWMENT FOR THE
Humanities



INSTITUTE of
Museum and Library
SERVICES

nature

Journal of
Political
Economy

JOURNAL OF
PUBLIC
ECONOMICS

Science

Developmental
Psychology

PLOS

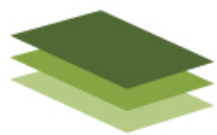
GORDON AND BETTY
MOORE
FOUNDATION

wellcometrust

It's required.

Data Management Plans

A data management plan (DMP) is a **written document** describing the **nature** and **structure** of the data you will likely use or produce in the course of research, along with your **strategies** for dealing with it **throughout and after** your project.



DMPTool

Guidance and Resources for your Data Management Plan

<https://dmptool.org/>

Common Elements of a DMP

1. Data description
2. Data documentation
3. Access, sharing, re-use
4. Storage and backups
5. Preservation and archiving
6. Resources and responsibilities



[flickr.com/photos/craightonmiller/8161895185](https://www.flickr.com/photos/craightonmiller/8161895185)

Collecting data



Photo by Jessica Trelogan

Test your plan

Automate where possible

Create snapshots

Ensure compliance

Office of Research Support:
research.utexas.edu/ors/

Re-using data

Find the right data

- Subject specialists: lib.utexas.edu/subject/index.php
- re3data.org

Integrate

Know your sources

- Restrictions
- Copyright
- Data citation

datacite.org

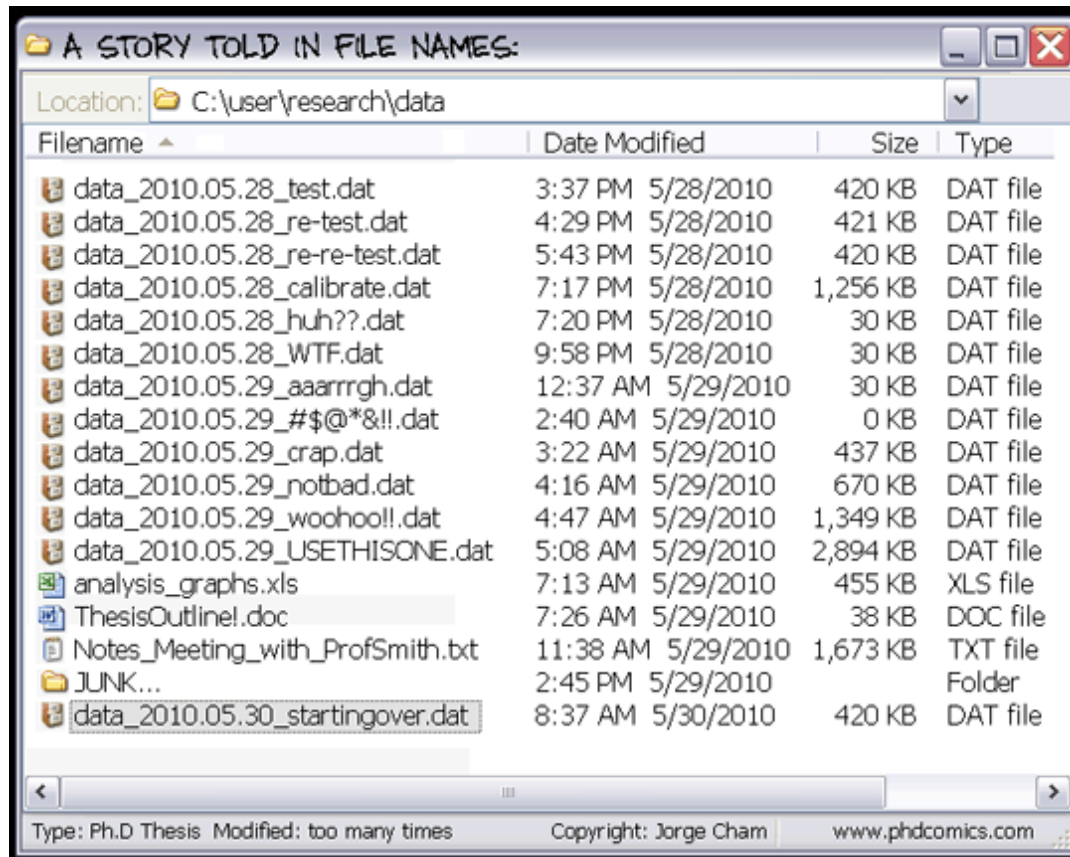
ailla.utexas.org/site/citation.html



Organize



File Names



<http://www.phdcomics.com/comics/archive.php?comid=1323>

Be descriptive, not generic

Include dates

CamelCase vs Pot_hole_case

No funny characters

"/ \ : * ? " < > [] & \$

Describe your convention

Use a batch re-namer

File Formats



Non-proprietary, open standards

Used commonly in your domain

Encoded with standard characters

Uncompressed (?)



[DROID](http://www.loc.gov/preservation/resources/rfs)

<http://www.loc.gov/preservation/resources/rfs>



Sharing active data



<https://www.tacc.utexas.edu/systems/stampede>

Ensure easy access

Avoid duplication

Control versions

Keep a list



Security

Passwords

Encryption

Updates

Backup strategies

Sensitive data

Storage options

University of Texas

- Departmental: server space by ATS
- 2 TB in Box
- UTMail (Google Drive)
- 5 TB at TACC
- ITS: VMs

Other Cloud

- DropBox
- Google Drive
- iCloud



By Evan-Amos - Own work, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=27940250>

Document, Document, Document



Photo courtesy of Institute of Classical Archaeology

Data only useful if understandable!

Metadata

Readme.txt (use a [template](#))

Codebooks/lab books/field notes

Data Dictionaries

Electronic Lab Notebooks

Managing Sensitive Data

- UT Information Security Office

Extended List of Confidential Data:

<https://security.utexas.edu/policies/extended-cat-1>

- Office of Research Support

Institutional Review Board (IRB)

Katharine Menke, IRB Program Coordinator

Institutional Animal Care and Use Committee (IACUC)

Institutional Biosafety Committee (IBC)

Conflict of Interest section

Preservation and Access

Texas ScholarWorks: to preserve and promote scholarly output. <https://repositories.lib.utexas.edu/>

re3data.org : find your disciplinary repository.

Archive of Indigenous Languages of Latin America:

<http://ailla.utexas.org/site/welcome.html>

UT Dataverse: (coming this Fall) to preserve and provide access to research data. <http://data.tdl.org/>



RESEARCH DATA SERVICES

Central location for accessing data management resources on campus

<http://lib.utexas.edu/datamanagement>

Jessica Trelogan

Data Management Coordinator

512-495-4267

j.trelogan@austin.utexas.edu

Upcoming Workshops

Introduction to Copyright & Fair Use

October 6, 2-3pm, PCL Learning Lab 2

Introduction to OpenRefine

October 17, 1-2:30pm, Scholars Commons at PCL, Data Lab

Writing a Data Management Plan

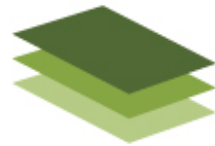
November 2, 12-1pm, PCL Learning Lab 2



DMP Deep Dive

Common Elements of a DMP

1. Data description



DMPTool

Guidance and Resources for your Data Management Plan

2. Data documentation

<https://dmptool.org/>

3. Access, sharing, and re-use

4. Storage and backups

5. Preservation and archiving

6. Resources and responsibilities

SPARC*

<http://datasharing.sparcopen.org/>

1. Data Description

What data will you gather or create?

- File types, formats, volume
- Methods and context of data collection
- Discussion of data sources
- Structure and organization of data files
- Data validation, quality assurance
- Data transformations or processing steps



<http://www.csr.utexas.edu/rs/gallery/valley/dscn0064.gif>

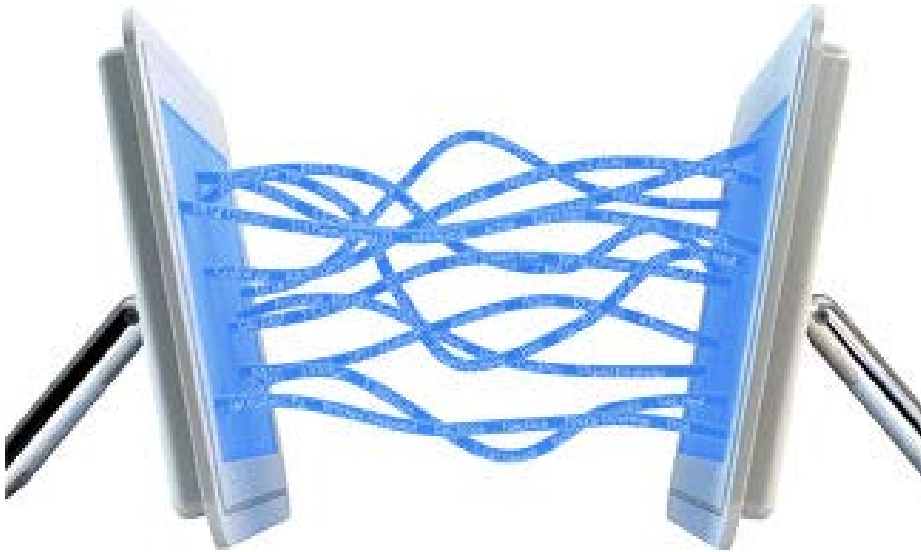
2. Metadata

What documentation will accompany your data?

- Type and form
- Metadata standards
- Basic details
- Definitions of variables, units, codes



3. Access, Sharing, and Reuse



From: <http://www.trendmls.com/guest/News/ShowDoc.aspx?id=771>

- Have you gained consent?
- Who will have access?
When? How?
- Are there any restrictions?
- What are the approved uses?
- How will you protect sensitive information?

4. Storage and Backups

Where will you store and back up your active data?

- Do you have enough storage space?
- Do you need security measures?
- How/how often will you do backups?
- What's your recovery plan?



https://c2.staticflickr.com/6/5304/5699142587_4d7b539a6c_b.jpg

5. Preservation and Archiving

What is your long-term preservation plan?



- What data should be retained?
Shared? Destroyed?
- How will you maintain and curate it?
- What future uses are there?
- Where will data live after the project? For how long?
- Are there any future costs?

6. Resources and Responsibilities

- Will you need additional help?
- Software? Hardware?
- What is this going to cost?
- Who is responsible for what?



<https://www.tacc.utexas.edu/systems/stampedo>